

**The pine bark adelgid *Pineus strobi*
contains two novel bacteriocyte-associated
gammaproteobacterial symbionts**

Elena R. Toenshoff, Gitta Szabó *et al.*

Supplementary information

Table S1. Populations of adelgids analyzed in this study

Organism	Location (year)	Host plant	GenBank accession numbers (insect host)	
			<i>col</i>	<i>eflalpha</i>
<i>Pineus strobi</i> (Hartig 1837)	Kaltern, Italy (2006)	<i>Pinus strobus</i> L. (eastern white pine)	KC784363	KC784365
<i>Pineus strobi</i> (Hartig 1837)	Germany, the origin of the host tree is unknown (2011)	<i>Pinus</i> sp.	KC784364	KC784366

Table S2. Primers and probes used in this study

Short name	Sequence (5' - 3')	Specificity	Target site (<i>E. coli</i>)	T _a / % FA ^a	Reference
Host gene targeting primers					
911	TTT CTA CAA ATC ATA AAG ATA TTG G	Mitochondrial cytochrome c oxidase subunit 1 (<i>col</i>), various eukaryotes	-	50	(1)
912	TAA ACT TCA GGG TGA CCA AAA AAT CA	Mitochondrial cytochrome c oxidase subunit 1 (<i>col</i>), various eukaryotes	-	50	(1)
AdeIEF1F1	GTA CAT CCC AAG CCG ATT GT	Partial nuclear elongation factor 1-alpha (<i>eflalpha</i>), various Hemiptera	-	61	(2)
AdeIEF1F2	CTC CAG CTA CAA AAC CAC GA	Partial nuclear elongation factor 1-alpha (<i>eflalpha</i>), various Hemiptera	-	61	(2)
16S and 23S rRNA primers					
616V	AGA GTT TGA TYM TGG CTC	16S rRNA gene, most Bacteria	8-25	52	(3)
1492R	GGY TAC CTT GTT ACG ACT T	16S rRNA gene, most Bacteria and Archaea	1492-1510	52	(4)
909F	ACT CAA AKG AAT WGA CCG	16S rRNA gene, most Bacteria	909-926	52	(5)
Ba1492R	NTA CCT TGT TAC GAC T	16S rRNA gene, most Bacteria and Archaea	1493-1508	52	(5)
AnnPi-1381R	TGG AAA CAT ATT CAC CGT G	16S rRNA gene, ' <i>Candidatus Annandia pinicola</i> '	1366-1384	60	this study
AnnPi-327F	TAC GGT CCA GAC TCT TAC	16S rRNA gene, ' <i>Candidatus Annandia pinicola</i> '	327-344	60	this study
255f	AGT AGY GGC GAG CGA A	23S rRNA gene, Bacteria	241-255	50	(6)
1037R	CGA CAA GGA ATT TCG CTA C	23S rRNA gene, Bacteria	1930-1948	50	(7)
Oligonucleotide probes					
EUB338-I	GCT GCC TCC CGT AGG AGT	16S rRNA gene, most Bacteria	338-355	10-50	(8)
HarPi-265	TCG TTG CCT AGG GGA GCC	16S rRNA gene, ' <i>Candidatus Hartigia pinicola</i> '	265-282	10-35	this study
AnnPi-327	GTA AGA GTC TGG ACC GTA	16S rRNA gene, ' <i>Candidatus Annandia pinicola</i> '	327-344	10-35	this study
HarPi-378	AAG ATA ACA CGT GCC CCG	23S rRNA gene, ' <i>Candidatus Hartigia pinicola</i> '	378-395	10-20	this study
AnnPi-1439	CGC TCT CTT GGC CAA CTT	23S rRNA gene, ' <i>Candidatus Annandia pinicola</i> '	1439-1456	10-35	this study

^a T_a, annealing temperature; % FA, formamide concentration



Figure S1. *Pineus strobi* (exules) sampled in Kaltern and feeding on the secondary host tree pine (*Pinus strobus*). The adelgids recognizable as white dots (grey rectangle) were located on the branch around the needles.

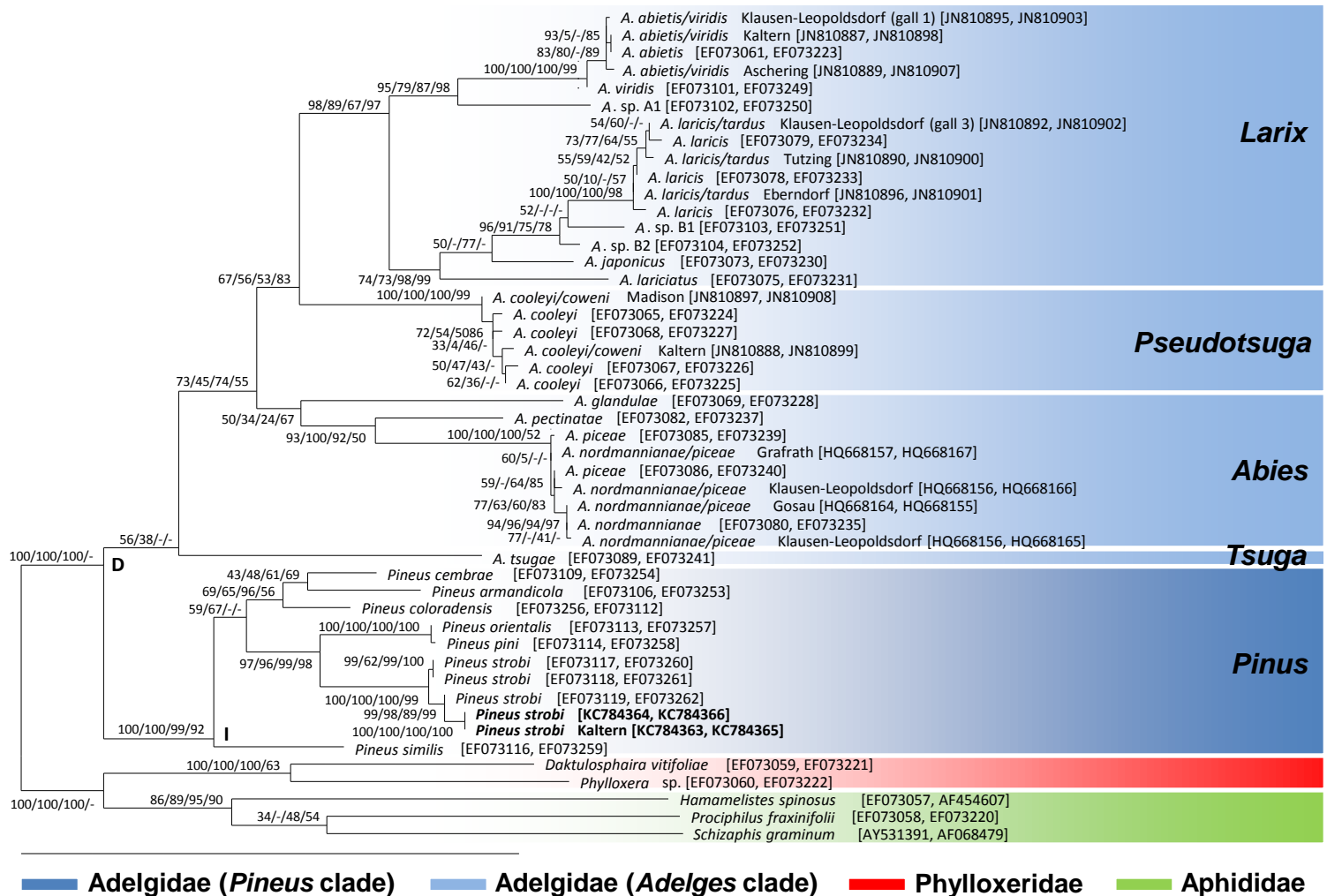


Figure S2. Phylogenetic affiliation of *Pineus strobi* with the insect family Adelgidae based on a concatenated dataset of the cytochrome c oxidase subunit 1 (*coI*, 624 nuc) and elongation factor 1-alpha (*ef1alpha*, 616 nuc) genes. Tree calculations were performed using maximum parsimony and TREEPUZZLE implemented in the ARB software package (9). PhyML trees were calculated using the Mobyly portal (<http://mobyly.pasteur.fr/cgi-bin/portal.py>; (10) and Neighbour-joining trees using MEGA (11). A maximum likelihood tree is shown. Maximum likelihood (1000 replicates), maximum parsimony (1000 replicates) and Neighbor-joining (1000 resamplings) bootstrap values, as well as TREEPUZZLE support values are indicated at the internal nodes. Selected members of the Aphididae and Phylloxeridae were used as outgroup. The letters (D = 88 ± 14.09 , I = 55 ± 11.67) indicate the estimated divergence times (millions of years \pm standard deviation) according to Havill *et al.* (12). GenBank/EMBL/DDBJ accession numbers of *coI* and *ef1alpha* are given in square brackets. Bar, 10% estimated evolutionary distance.

References

1. **Guryev V, Makarevitch I, Blinov A, Martin J. 2001.** Phylogeny of the genus *Chironomus* (Diptera) inferred from DNA sequences of mitochondrial cytochrome b and cytochrome oxidase I. *Mol. Phylogenet. Evol.* **19**:9–21.
2. **Havill NP, Foottit RG, von Dohlen CD. 2007.** Evolution of host specialization in the Adelgidae (Insecta: Hemiptera) inferred from molecular phylogenetics. *Mol. Phylogenet. Evol.* **44**:357–370.
3. **Juretschko S, Timmermann G, Schmid M, Schleifer K-H, Pommerening-Röser A, Koops H-P, Wagner M. 1998.** Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and *Nitrospira*-like bacteria as dominant populations. *Appl. Environ. Microbiol.* **64**:3042–3051.
4. **Loy A, Schulz C, Lücker S, Schöpfer-Wendels A, Stoecker K, Baranyi C, Lehner A, Wagner M. 2005.** 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order “Rhodocyclales”. *Appl. Environ. Microbiol.* **71**:1373–1386.
5. **Berry D, Mahfoudh KB, Wagner M, Loy A. 2011.** Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Appl. Environ. Microbiol.* **77**:7846–7849.
6. **Lane DJ. 1991.** 16S/23S rRNA sequencing. in *Nucleic acid techniques in bacterial systematics*. eds Stackebrandt E., Goodfellow M. (John Wiley & Sons, Chichester, United Kingdom)
7. **Ludwig W, Kirchhof G, Klugbauer N, Weizenegger M, Betzl D, Ehrmann M, Hertel C, Jilg S, Tatzel R, Zitzelsberger H, Liebl S, Hochberger M, Shah J, Lane D, Wallnöfer PR, Scheifer KH. 1992.** Complete 23S ribosomal RNA sequences of Gram-positive bacteria with a low DNA G+C content. *Syst. Appl. Microbiol.* **15**:487–501.
8. **Amann RI, Binder BJ, Olson RJ, Chisholm SW, Devereux R, Stahl DA. 1990.** Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**:1919–1925.
9. **Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüßmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer K-H. 2004.** ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363–1371.
10. **Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010.** New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**:307–321.
11. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011.** MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**:2731–2739.
12. **Havill NP, Foottit RG, von Dohlen CD. 2007.** Evolution of host specialization in the Adelgidae (Insecta: Hemiptera) inferred from molecular phylogenetics. *Mol. Phylogenet. Evol.* **44**:357–370.